RECIIS
Electronic Journal of Communication Information & Innovation in Health
[www.reciis.cict.fiocruz.br]
ISSN 1981-6286

Editorial

Ontologies, semantic Web and health

DOI: 10.3395/reciis.v3i1.238en

Frederico Freitas
Informatics Center, Pernambuco Federal University, Recife, Brazil
fred@cin.ufpe.br

Stefan Schulz
Institute of Medical Biometry and Medical Informatics, University Medical Center Freiburg, Germany
stschulz@uni-freiburg.de

With a long tradition of study in the philosophy area, going back at least 23 centuries, the term "Ontology" has become one of the most fashionable terms in today's information technology world, being applied to systems in many other fields, with the main ones including biology and medicine. While the term "Ontology" used to denote a branch of metaphysics, nowadays "ontologies" are understood as formal vocabularies that describes the basic premises of a certain domain.

There is at least one main reason for all this interest on behalf of computer scientists. According to Tim Bernes-Lee, one of the main people responsible for consolidating the internet itself, ontologies form the main component and motivation of the semantic Web, a Web where programs and systems are capable of "understanding and processing" page information, according to the context. But what are ontologies and how do they help systems to process data with such depth?

In practical words, ontologies standardize meaning through semantic identifiers, which can represent the real and conceptual worlds. Ontologies form definitions of concepts, classes, properties, relationships, restrictions, and axioms on a certain domain (for example, neurology).

In the articles of this thematic issue, many examples of ontologies are presented, varying in complexity, expressiveness, etc. So, if a Web page references one or more ontologies (for example, if a fragment of the page is a professor's name, referencing this concept with an ontology about Academia), it is attributing meaning to the content of the page so that software can use other definitions of the ontology which define the *a priori* relationships and restrictions related to the referenced concept and then interpret other parts of the page within the context defined by the ontology. For example, when it finds a list of doctorate students on the professor's page, the software can interpret that those students are oriented by the professor, and work in areas researched by the professor (and which may also be on the professor's page), and are enrolled in the post-graduate program the professor belongs to. It is worth noting that this information may not be literally present on the Web page. Therefore, processing information using ontologies, which provide an excellent context for understanding the information, both for human users as for software agents, is becoming a trend in various areas and types of application. The semantic Web, for example, should have a strong impact on the Web's commercial locations; ontologies can also be used as vocabulary in a message exchange between the so-called intelligent agents, software entities that can reason on knowledge, allowing these agents to negotiate purchase requests on behalf of their companies – and there are already academic prototypes simulating this situation.

Among the most common areas of ontology application are medicine and biology, with some centers exclusively dedicated to the study of this technology. In the United States, the Stanford Center for Biomedical Informatics Research created the most used ontologies editor, the Protégé. There is also a virtual organization, the National Center for Biomedical Ontologies, which involves the research groups from Stanford, Victoria and Buffalo universities, as well as the Mayo Clinic in Rochester. In Europe, the Ontologies in Medicine Research Group at the Institute for Medical Informatics, Statistics and Epidemiology in Leipzig, and the Health Informatics Research Group at the University Medical Center in Freiburg, both in Germany, and the Bio-Health Informatics Group, at the University of Manchester, in England, are among those doing the most research in biological ontologies and the semantic Web. The complexity of medical and biological knowledge makes it difficult to put together traditional systems, for in order to assist medical tasks, the systems need much knowledge and inference capacity. This may be the main reason for the applicability and subsequent success of ontology use in these areas. We discussed these needs a bit more in the first article. It must simply be emphasized that the volume of research produced linking ontologies, the semantic Web and health justifies the launch of this thematic issue and we accepted the task with the conviction that RECIIS is, in fact, a vehicle where all these research topics are contemplated. The community that researches these topics answered the call for works satisfactorily and the publishers are grateful for the contributions that were sent to us.

The reviewers selected five articles out of the submissions received for this thematic issue. Along with this introductory article from the guest publishers, they make up the six articles in this thematic issue. When putting together the supplement, we tried to order the articles so as to gradually introduce the concepts and practices in biomedical ontologies in an ascending order of complexity, as the first two themes of the thematic issue – Ontologies and the Semantic Web – are not yet widespread in Brazil. Below, we will briefly describe each of the articles in the supplement.

The first article, entitled “Survey of the current terminologies and ontologies in biology and medicine” by one of the guest publishers, aims to present introductory concepts about ontologies and terminological systems in biology and medicine, giving a broad view of the subject with the description of more widespread and/or emblematic systems. After a brief discussion of the use of terminologies in contrast with the use of ontologies, the systems ICD (International Classification of Diseases), SNOMED CT (Systematized Nomenclature of Medicine – Clinical Terms), MeSH (Medical Subject Headings), OpenGALEN (General Architecture for Languages, Encyclopaedias and Nomenclatures in Medicine), FMA (Foundational Model of Anatomy), as well as the initiatives UMLS (Unified Medical Language System), the OBO Foundry (Open Biomedical Ontologies Foundry) in general and the GO (Gene Ontology) in particular.

The article “Ontological and conceptual bases for a scientific knowledge model in biomedical articles”, written by Carlos Marcondes and co-authors, proposes a classification of scientific articles in the medicine area. Each class has a semantic annotation model, annotations which must be made using ontologies, so that the knowledge expressed in the articles can be processed and “understood” by computer systems. Before the proposal, the article brings a discussion which justifies the existence of the model based on scientific methodology and science philosophy and on the analysis of 75 medical articles. The model makes semantic recovery of article information possible, allowing searches through concepts and relationships, instead of key words as in traditional search engines.

The article “Advantages and limitations of formal ontologies in the biomedical domain”, by Stefan Schulz and co-authors (including Barry Smith, philosopher and author of various books and well-known articles about bio-ontologies and director of various organizations related to the subject, such as the OBO Foundry itself), serves as an excellent introduction to the mathematical formalisms used in biomedical ontologies, proposing a series of criteria to delimit the concept of knowledge representation resources ontology more broadly. The advantages and disadvantages of each representation are carefully discussed, from thesauruses, frequently available information sources in the biomedical area, to the expressive logic of descriptions – the most expressive knowledge representation formalism standardized by the World Wide Web Consortium (W3C). The discussion is mainly around the cost-benefit relationship of adding expressiveness to the form of representation in use. Simple examples using description logic are introduced, in gradual complexity, thus helping the reader to have a concrete idea of how complex biomedical concepts are represented and how these concepts are used during the automatic reasoning process, as well as possible modeling slips that may lead to incorrect automatic reasoning.

The next article, “An ontological analysis of the electrocardiogram”, by Bernardo Gonçalves and co-authors, describes an ontology that has been carefully constructed to allow reasoning on a complex, specific and challenging domain in terms of representation, the performance of an Eletrocardiogram. The ontology was elaborated using two fairly rich and popular biomedical ontologies, the OBO and the FMA – already described in the first article – and the UFO (Unified Foundational Ontology) top ontology. This top ontology lends solid bases to the representation of complex elements directly or indirectly present in electrocardiograms, such as part-whole relationships and temporal production, characterization and generation relationships, classifying them with their appropriate meta-characteristics, whether these are symmetrical or anti-symmetrical, reflexive or irreflexive or transitive or intransitive. Therefore, the elaboration of the ECG ontology is a good example of how to employ well-founded modeling principles to guarantee that the automatic reasoner will not draw erroneous conclusions due to a lack of basic knowledge, such as the consequences of one concept being part of another. At the end of the article, some ways to apply the knowledge provided are suggested.

The last two articles are dedicated to the construction of ontologies, but not from a modeling point of view. The article “Methodological Aspects in Ontology Reuse: a study based on the genomic annotations in the

Trypanosomatides domain," by Maria Luiza Campos and co-authors, discusses the reuse of ontologies in general lines, also describing their experiments in the biomedical area, the description of the Trypanosomatides genome. An aligner was employed, with the function of comparing two ontologies, showing the correspondences between the elements of these ontologies, a task correlated to the reuse task. The authors end the article with a list of conclusions about their specific experiment, that is, in what cases the aligner corresponded to the authors' expectations.

Finally, the article "Automatic extraction of composite terms for ontology construction: An experiment in the health area," by Lucelene Lopes and Renata Vieira, discusses an automatic learning tool for ontologies, elaborated within the authors' research group. The tool is founded on specific techniques of a specific study area of artificial intelligence, known as Natural Language Processing, which is based on text processing from the linguistic point of view, generating, therefore, a much more profound processing of textual information. These techniques are used to extract elements that will make up an ontology. This type of approach is particularly interesting as it automates the knowledge production process for ontologies, accelerating the process of obtaining them. The authors describe initial experiments with corpus on pediatrics. The experiments showed that approaches based on learning can indeed be useful in the construction phase of terminology for an ontology, however, they do not guarantee the coverage necessary to include all the terminology contained in the corpus.

For our final words, we would like to thank the reviewers, Guilherme Ataíde, Werner Ceusters, Ronald Cornet, Marcos Galindo, Rosario Girardi, Giancarlo Guizzardi, Robert Hoehndorf and César Tacla, for a serious and high quality review job, without which this issue probably could not have been produced. The guest publishers would also like to thank the Scientific Editors of RECIIS, Carlos Saldanha Machado, and Josué Laguardia for guiding us during the whole publishing process, with several useful tips, being very attentive and promptly answering a long list of questions that arose during this work, as well as presenting us with the opportunity to publish it.

About the editors

Frederico Freitas

Received his PhD at the University of Santa Catarina, Brazil, and is currently affiliated to the Center for Informatics at the Federal University of Pernambuco, Brazil (CIn/UFPE). He researched for almost a year at the Department of Informatics of the University of Karlsruhe, as a member of the Brazilian-German project "A semantic approach to data retrieval". He published a number of papers at renowned conferences and workshops such as IJCAI and other sponsored by ACM (Association on Computer Machinery) and IEEE (Institute of Electrical and Electronical Engineering). He has also co-chaired two workshop series: WONTO (Workshop on Ontologies and their Applications) in Brazil and BAOSW (Building Applications with Ontologies for the Semantic Web) in Portugal. He co-edited Special Issues on related themes at JBCS (Journal of Brazilian Computer Society) and JUCS (Journal of Universal Computer Science). He currently has collaborations with the University of Paul Cezanne at Marseille and INRIA, Montbonnot, in France and the Universities of Karlsruhe, Freiburg and Mannheim in Germany. His interest areas comprise ontologies, multi-agent systems, knowledge representation, mediation and text mining.

Stefan Schulz

Holds a medical degree (Heidelberg University, Germany) and is senior researcher and professor at the Institute for Medical Biometry and Medical Informatics of the University Medical Center Freiburg, where he leads the Medical Informatics Research Group. His work focuses on biomedical terminologies and ontologies, biomedical knowledge representation, cross-language medical document retrieval, text and data mining in clinical document repositories, eLearning in Medicine, and health informatics in developing countries.

After clinical work in surgery and internal medicine he obtained his doctoral degree in the field of tropical hygiene where he carried out a parasitological field study on in São Luís, Brazil. After obtaining a technical qualification in medical computing, he moved to the University of Freiburg, where he participated in clinical and educational software development projects and participated in several research projects in the field of information extraction, biomedical terminologies, medical language engineering and semantic technologies. He have been played a leading role in severe EU-funded research projects. Stefan Schulz is author of more than hundred peer reviewed publications and has received several awards. Since 2001 he has repeatedly contributed to Brazilian health informatics research projects as a visiting researcher at the Paraná Catholic University (PUC-PR).